The variability in the effect of additional education on different mathematics skills in primary school - A regression discontinuity analysis

Bartelet, D.; Haelermans, C.; Groot, W.N.J.; Maassen van den Brink, H.

The variability in the effect

of additional education on different mathematical skills

in primary school - A regression discontinuity analysis

Dimona Bartelet, Carla Haelermans,

Wim Groot & Henriëtte Maassen van den Brink

TIER WORKING PAPER SERIES
TIER WP 13/12

Universiteit van Amsterdam Maastricht University rijksuniversiteit groningen

The variability in the effect of additional education on different mathematical skills in primary school - A regression discontinuity analysis

Dimona Bartelet[1,2,*], Carla Haelermans[1,3], Wim Groot[1] and Henriëtte Maassen van den Brink[1]

[1] *Top Institute for Evidence Based Education Research, Maastricht University, Maastricht 6200 MD, The Netherlands*

[2] *Faculty of Psychology & Neuroscience, Maastricht University, Maastricht 6200 MD, The Netherlands*

[3] *Centre for Innovations and Public Sector Efficiency Studies, Delft University of Technology, The Netherlands*

Abstract

This paper explores the variability in the effect of an additional year of education on different basic mathematical skills, which are taught to children and explicitly repeated at different points in time during elementary school. In addition, the role of child specific characteristics and the role of the school is addressed. Using a regression discontinuity approach, we estimate the local education effect for a range of successive elementary school grade cohorts at six schools in the Netherlands, while including child specific characteristics and school fixed effects. The results indicate that the effect of education is contingent upon the mathematical skills and grade levels studied. A significant education effect is observed the year after children have first learned a mathematical skill and in the last year of primary school, which can be explained by the additional focus that is being put on elementary mathematics skills in these years. Furthermore, the outcomes suggest that children's multiplication skills are significantly lower if their school emphasizes child centered, natural learning and multigrade teaching instead of traditional classroom learning.

JEL-Classification – I20, I21.

Key words – return to education; mathematic abilities; primary school, regression discontinuity.

**Corresponding author at: P.O. Box 616, 6200 MD Maastricht, The Netherlands*

E-mail: dimona.bartelet@maastrichtuniversity.nl

Phone: +31 43 3884698

1. Introduction

The development of proficient mathematics skills is deemed important for an individual's daily life and future success in an increasingly quantitative world (Dougherty, 2003; Murnane, Willett, & Levy, 1995). Developmental theory states that more advanced mathematics skills are built upon more elementary ones (Entwisle, Alexander, & Olson, 2005). In addition, elementary mathematics skills have been linked to adulthood socioeconomic status and earnings (Crawford & Cribb, 2013; Ritchie & Bates, in press).

Although some basic mathematics skills are believed to be inherent, many elementary mathematics skills are proposed to be developed only through formal schooling and not merely through maturational factors such as age (Geary, 1995). Accordingly, many ministries of education in western societies consider the development of mathematic skills primarily a responsibility of schools (e.g. Rijksoverheid, 2013a). Yet, discussions concerning children's insufficiently developed mathematics skills (van Groenestijn, 2008), have raised questions about the actual effectiveness of education in primary school grades. In the Netherlands, the Ministry of Education has formulated goals outlining at the end of which grade level of primary school children should master certain skills in mathematics (van der Stap, 2012). In consecutive grades children are expected to improve these skills. Underlying this expectation is the assumption that an additional year of education leads to a higher proficiency in mathematics skills. However Luyten (2006) who investigated the effect of one additional year of education received by 4th grade children compared to 3rd graders did not consistently observe a significant positive education effect on mathematics achievement across different countries. This leads to the question of whether the investment in an additional year of education does in fact produce significant and linear advancements in children's mathematics skills, which they are expected to acquire through formal education, beyond the effect of maturational factors.

An extensive amount of research estimating the effect of an additional year of education have reported a significant effect, but most of these studies focused on long-term effects such as wage and social capital (Harmon, Oosterbeek, & Walker, 2003; Huang, Maassen van den Brink, & Groot, 2009). Despite the importance of knowing the impact of an extra year of education on the development of elementary mathematics skills in primary school, only few studies addressed this issue (e.g. Cahan & Davis, 1987; Cliffordson, 2010). These studies mainly used general mathematics achievement tests as outcome measure. Yet, since mathematics is a complex subject area consisting of multiple sub-skills, which are taught and repeated at different points in time during primary school (e.g. van der Stap, 2012; von Aster

& Shalev, 2007), it might be fallacious to generalize results obtained by a composite measure of mathematics ability to different sub-skills and grade levels. Also, previous studies rarely controlled for other cognitive child-specific characteristics such as general response speed and reading ability when estimating the return to education on mathematics achievement.

Hence, we pose the question what the effect is of an additional year of education on various mathematics sub-skills in different grade levels and to what extent child-specific background characteristics and school differences play a role. In order to answer this question we implement a regression discontinuity (RD) design using age in months at testing moment as the forcing variable and a bandwith of 3 months. T-tests and Mann-Whitney tests on different background variables indicated that the control and treatment groups are equal on several observable characteristics. Analyzing micro-level data of 1200 grade 2 through grade 6 children, we find that the presence of an education effect is dependent upon the mathematics sub-skill and grade cohorts assessed. Altogether, a significant impact is observed in the youngest and oldest grade cohorts, which corresponds to the grade levels in which children have just learned the mathematics sub-skills and explicitly repeat these sub-skills in class, respectively. The magnitude of the education effect is similar for both grade cohorts. Lastly, school is found to play an important role in the development of one mathematic sub-skill, namely mental arithmetic. Specifically, children enrolled in the school emphasizing a child centered, natural learning and multigrade teaching perform significantly worse than children attending schools characterized by a more traditional learning and teaching approach.

The contribution of this paper to the literature is twofold: First, we are able to determine the consistency of the local education effect across multiple elementary mathematics sub-skills and a wide range of primary school grades. Second, child-specific characteristics and school fixed effects are included in the specifications, allowing for a more precise and accurate estimation of the returns to education. Furthermore, it enables us to address the influence of child-specific characteristics and school effects on mathematics skills, while controlling for the effect of education. With respect to school effects, we are particularly interested in whether performance differences between children attending the reference school, which emphasizes child centered, natural learning and multigrade teaching, and children enrolled in the other sampled schools, which use a more traditional teaching approach, remain after controlling for endogenous sorting of students across schools and the amount of education received.

In the remainder of this paper we first briefly review the literature on the effect of education on pupil’s mathematics achievement. Subsequently, in Section 3, we discuss the data set, whereas in Section 4 we describe the methodology and identification strategy. The RD results are presented in Section 5. Section 6 concludes this paper and discusses the results in light of the current literature.

2. Literature overview

Studies increasingly use quasi-experimental methods to estimate the effect of education on mathematics achievement in primary school. In the upcoming literature overview, we chiefly focus on studies that have investigated the return to one additional year of education in primary school. One of the major challenges in specifying the effect of additional education is its strong correlation with age. In western nations older children have generally received more education than younger ones due to school entry laws. To disentangle the causal effect of age and schooling on mathematics achievement, Cahan and Davis (1987) applied the fuzzy RD approach. They attributed two-third of the observed mathematics achievement differences between grade 1 and grade 2 children to a one year difference in amount of education received. Since then, similar findings have been reported for different grade levels and countries (e.g. Cliffordson, 2010; Luyten, 2006).

However, Cliffordson (2010) noticed that the education effect between 7^{th} and 8^{th} grade was smaller than between 6^{th} and 7^{th} grade. According to the author this discrepancy might be due to a negative relationship between children's age and the effect of education. Moreover, the author mentioned that the relative small sample size of the 8^{th} grade cohort compared to the other two cohorts, could have led the 8^{th} grade cohort to be more prone to random effects. In line with the finding by Cliffordson (2010), Marcotte (2007) found that a decrease in the quantity of schooling due to heavy snowfall, affected the mathematics achievement of grade 3 pupils more than of grade 5 and grade 8 children.

Also, two studies comparing kindergarteners and grade 1 children examined if the education effect is consistent across elementary sub-skills of mathematics. Bisanz, Dunn, and Morrison (1995) measured children's mental arithmetic accuracy and conservation of number abilities. While little explicit attention is directed to the latter skill, teaching children mental arithmetic abilities is a key goal in the early years of formal education. As hypothesized, Bisanz and colleagues (1995) observed an education effect on arithmetic abilities, while children's conservation of number skills only improved as a function of age. In a study by Naito and Miura (2001), an addition task and numerous number concept tasks were administered. Children's place-value understanding, but not their number conservation skills and addition accuracy was not a function of schooling.

Some progress has also been made towards identifying child-level and school-level characteristics which affect the impact of education on mathematics abilities. Most consistent was the finding that the size of the education effects differs for schools (e.g. Heck & Moriyama, 2010; Luyten, 2006). Data of 4^{th} grade children analyzed by Sims (2008) also suggests that the impact of education differs in urban and rural districts, though he notes that the sample size was too small to formulate definite conclusions. Regarding child-level characteristics, Fitzpatrick, Grissmer, and Hastedt (2011) observed minor differences in the effect of an extra day of education between test moments for groups differing with respect to gender and race. Heck and Moriyama (2010) did not find gender to significantly interact with education. However, Luyten (2006) found girl's performance to be more affected by education than boy's mathematic achievement in Iceland and Scotland, but in none of the other countries included, indicating that whether gender moderates the return to education might be country dependent. Lastly, researchers note that children's intellectual ability, which has been related to children's mathematic achievement (Lu, Weber, & Spinath, 2011; Spinath, Freudenthaler, & Neubauer, 2010), should be taken into account when estimating return to education on school achievement (e.g. Gottfried, 2010; Heckman & Vytlacil, 2001).

To recapitulate, the results of previous studies generally report a positive effect of one extra year of education on mathematics achievement for different primary school grade levels, indicating that formal education does lead to attainment gains beyond improvements attributable to maturational factors. Moreover, studies find the effect of education to vary depending on child-level and school-level characteristics, supporting the importance of including child specific characteristics and school fixed effects in analyses aimed at estimating education effects.

3. Data description

Participants

In the context of a test standardization project, cross-sectional data of 1200 second through sixth grade children, from six schools located across the Netherlands, was gathered. Child-level variables were collected during individual testing sessions, lasting about an hour, from January through June 2011. Except for the nonverbal intelligence (IQ) task, the cognitive measures were computerized. Characteristics of children's schools were derived from an existing dataset from the Dutch Ministry of Education[1]. The summary statistics for

[1] DUO (Dienst Uitvoering Onderwijs)

the sample are presented in Table 1, and they are discussed below for the different variables separately.

Table 1 - Summary statistics complete sample (N = 1200)

	Measurement unit	Observations	Mean	St. Dev.	Minimum	Maximum
Outcome variable						
mental arithmetic	acc-RT	1197	0,57	0,38	0,03	2,13
ordinality	acc-RT	1198	0,44	0,17	0,05	1,08
number line 0-1000	absolute error	1171	9,49	7,06	1,63	43,55
Child-specific characteristics						
age	months	1200	120,36	17,41	82,00	162,00
gender	1 = boys; 0 = girls	1200	0,49	0,50	0,00	1,00
mother tongue	1 = dutch; 0 = bilingual	1200	0,91	0,29	0,00	1,00
nonverbal IQ	items correct	1200	30,19	3,68	16,00	36,00
reading ability	items correct	1196	118,67	26,52	23,00	191,00
School characteristics						
ftes 2010	total	1200	24,59	8,63	15,10	37,00
female teachers	percent	1200	0,87	0,06	0,79	0,95
weight pupils population	percent	1200	0,07	0,04	0,02	0,15
number of pupils	pupils	1200	455,72	145,37	293,00	675,00

Treatment effect

The treatment of interest is one additional year of education, operationalized as being enrolled in one grade level higher than the control group. Given that the data set comprises children from multiple successive grade levels who all finished the same mathematics skill tasks, models can be estimated for four different grade cohorts consisting of two grade levels. These four different grade cohorts are grades 2 and 3, grades 3 and 4, grades 4 and 5, and grades 5 and 6.

Outcome measures

Three elementary sub-skills of mathematics, which children are expected to learn in primary school were assessed, namely mental arithmetic, number line placement and ordinal judgment. Mental arithmetic skills are operationalized using a task in which multiplication sums are displayed together with two possible outcomes. Children were instructed to choose the correct answer as fast as possible. Multiplication sums are first taught in second grade and by the mid of third grade children should have mastered all multiplication tables from 1 to 10 (van der Stap, 2012). Another aim of the early math school curriculum is to teach children the ordinal property of numbers. Specifically, children should understand by the end of second grade (van der Stap, 2012) that numbers from 0 to 100 do not only tell you the quantity of something, but also indicate which position an item has in an ordered sequence

(Jacob & Nieder, 2008). This skill is assessed with the ordinality task in which a string of three one-digit or two-digit numbers were presented to children, followed by the question whether the numbers were displayed in the correct order from left to right. Lastly, children are supposed to have acquired the ability to place numbers from 0 to 1000 correctly on a visually presented number line by mid third grade (van der Stap, 2012). This skill was examined using a number line task which had a starting point marked by the digit zero and an endpoint labeled by the digit 1000.

Table 1 also presents the summary statistics for these tasks over the entire sample. On average, children were able to correctly solve 0.572 mental arithmetic problems per second. However, children's mental arithmetic capacity varies greatly, as some children answered as few as 0.030 correct per second, while others answered about two correct per second. Although the variation in ordinal judgment performance is less, children's score on this sub-skill of mathematics still ranges from correctly answering 0.050 stimuli per second to 1.080 stimuli per second. Their mean performance on this task is 0.437 correct answers per second. Children's number line placement skill was quantified using percent of absolute error scores. Also on this task substantial differences exist among children as indicated by an observed minimum score of 1.630 and a maximum score of 43.550. The average score for this outcome measure is 9.485 percent absolute errors. In 28 cases, of which 93 percent was enrolled in grade 2, the number line task was not administered due to insufficient knowledge of three-digit Arabic numbers.

Control variables

The dataset also contains information on children's age, gender, mother tongue, school and nonverbal IQ, allowing us to examine the impact of child specific characteristics and school fixed effects.

Table 1 shows that of the total sample included in this study, 51 percent are girls (coded 0) and 49 percent boys (coded 1), whose age ranges from 83 months (6;9 years) to 165 months (13;8 years). The mean age is 121 (10;1 years) months with a standard deviation of 17.297 months (1;4 years). Furthermore, 110 children report having a different mother tongue than Dutch, speaking one of 27 different languages. The vast majority of these children are raised bilingual (N=107). Due to the low number of observations in the non-Dutch not bilingual group, the mother tongue variable is converted into a binary categorical variable by combining children with another mother tongue than Dutch into one group (coded 0).

Nonverbal IQ is assessed with the Coloured Progressive Matrices test (Raven et al., 1995), in which children see a colored pattern and are asked to select the missing piece out of 6 choices. This task is assumed to measure children's general aptitude to process nonverbal information, to analyze and solve ambiguous situations and their general capacity to learn (Bernstein, Penner, Clarke-Stewart & Roy, 2006). Children answered on average 30.191 of 36 items correctly. This corresponds to a score of 6.612 out of 10. In the reading ability task children are asked to read as many words as possible, increasing in difficulty, in 90 seconds. The maximum attainable score is 225 items read correctly, but the maximum obtained score is only 191 and the mean performance is 119 items.

Children attend one of six schools which are localized in four different regions of the Netherlands. Except for one school, which constitutes 11 percent of the sample, schools are about equally presented in the data set, ranging from 15 to 21 percent. Two of the schools are situated in a rural area, two in a medium populated urban area and two in a densely populated urban area. Furthermore, schools differ in the amount of pupils that are enrolled and their denomination, which is the primary religious believe of the school. In the Netherlands this is predominantly public (no specific religious believe), Catholic or Protestant. In our sample all denominations are equally represented. Moreover, the percentage of female teachers employed at the schools ranges from 79% to 95%, but is as expected overall high. One of the sampled schools distinguishes itself from the remaining schools (with traditional classroom learning) in emphasizing child centered, natural learning and multigrade teaching. Specifically, they believe that every child should be taught in a manner most suitable to the child's learning style and they emphasize self-reliance and collaboration. A child has a central role in determining its learning route whereby their interests are more important than their age (J.H. Snijdersschool, 2013). In addition, two to three different grade levels are educated within the same group (Inspectie van het Onderwijs, 2012). For the purpose of analysis, this school is treated as the reference group in the school fixed effect analyses.

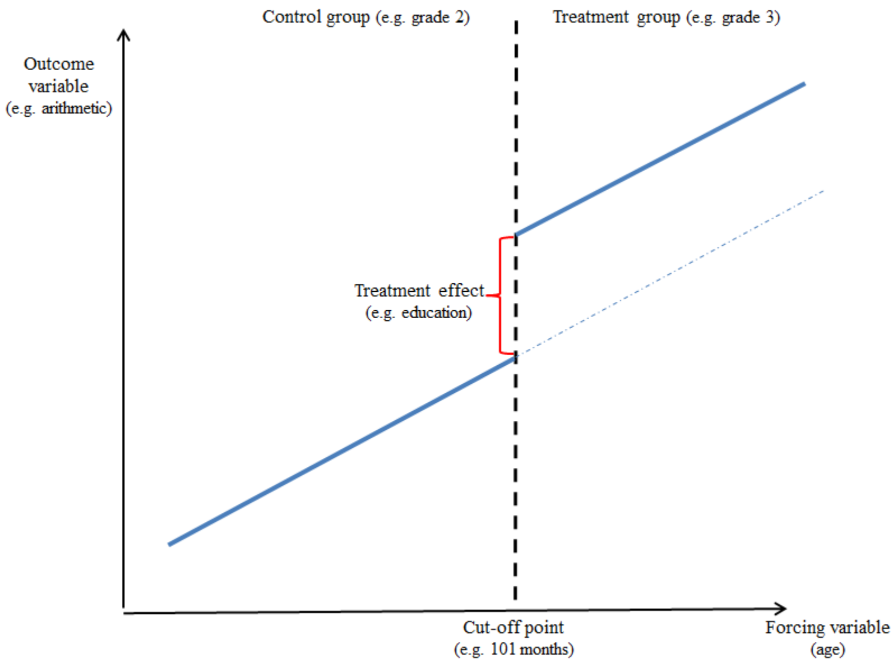
4. Methodology

Regression Discontinuity approach

The effect of an additional year of education on mathematics sub-skills for different primary school grade cohorts is estimated using the regression discontinuity method (RD) (Angrist & Pischke, 2009; Cook, 2008; Jacob & Lefgren, 2004; Luyten, 2006; Nichols, 2009; Schochet et al., 2010). A key prerequisite of this natural experiment approach is that assignment to the control and treatment depends upon individuals' scores on a so called forcing variable. Children who have a value above an exogenously chosen cut-off point are part of the treatment group, whereas the children who have a similar value below the cut-off point constitute the control group. If the assumption holds that individuals cannot precisely control on which side of the cut-off point they fall, then randomized variation exists around the cut-off point (Lee & Lemieux, 2009; Murnane & Willett, 2011). Thus, besides the forcing variable only idiosyncractic characteristics of individuals influence their probability of receiving a treatment. Accordingly, children who have a value on the forcing variable just below the cut-off are postulated to provide a counterfactual to the children scoring just above the cut-off and thus receiving the treatment (Lee & Lemieux, 2009). Another implication is that any observed discontinuity in the scores on the outcome measures between children scoring just below and those scoring just above the cut-off point can be attributed to the treatment. However, the derived RD estimate of the treatment effect is specific to the chosen cut-off point and therefore should be interpreted as a local average treatment effect. This means that the results are strictly speaking only applicable to individuals who have a forcing variable score equaling the cut-off value. This reduces the external validity of the results substantially, as they cannot be generalized across the entire population. Nevertheless, the strong internal validity of the RD method does provide us with valuable information on the treatment effect for individuals close the cut-off point.

Figure 1 illustrates the basic idea of a simple linear RD design graphically. The forcing variable is represented on the horizontal axis and the outcome variable on the solid vertical axis. The dashed horizontal line demarcating the horizontal axis designates the cut-off point. The sloping line on the left side of the cut-off point represents control children’s average scores on the outcome measure at different values of the forcing variable, while treatment children’s average scores are displayed by the sloping line on the right side of the cut-off point. At the cut-off point a discontinuity can be observed between the two sloping lines. The corresponding vertical distance reflects the causal effect of the treatment. Figure 1 also demonstrates a key requirement of the RD design for the results to be valid. Since the estimation of a treatment effect at the cut-off also relies on data points further away from the cut-off, choosing the correct functional form of the sloping lines is crucial. Another necessary assumption is that except for the treatment variable, all variables explaining the outcome measure exhibit no discontinuity at the cut-off value. In other words, the relation between these variables and the outcome measure should be smooth.

Figure 1 - Visual representation simple linear regression discontinuity design



A misspecification of the relationship between the forcing variable and the outcome measure leads to a biased estimate of the treatment effect. Choosing only observations close to the cut-off point to estimate the treatment effect, thus employing a small bandwith, has the advantage that the trend between the forcing and outcome variable is linear across the entire bandwith. If one selects a larger bandwith, it is increasingly important to test different functional forms in order to select the function which correctly models the relationship. However, the inclusion of more observations enhances the statistical power and precision of the RD design. Yet, this is traded off by a decrease in the credibility of the equality-in-expectation assumptions, affecting the studies internal validity. Consequently, it is critical to check this assumption by comparing the control and treatment group on several observable covariates.

Simple linear RD designs are modeled by the following equation:

$$Y_i = \beta_0 + \beta_1(x_i - x_c) + \beta_2 T + \varepsilon_i \quad (1)$$

where the regression coefficients β_1 and β_2 denote the forcing variable and treatment effect respectively. The parameter β_0 denotes the intercept at the cut-off point. The intercept has been recentered as a result of subtracting the cut-off value from the forcing variable value. This computation is denoted in the model by the expression $x_i - x_c$. For example, if age is the forcing variable and 101 months the chosen cut-off point, then x_i is substituted by an individual's age at the moment of testing and x_c by the cut-off value, which is 101 months.

Lastly, T stands for the dummy variable treatment (T=1: treatment) and ε_i is an error term. Note that in this model the effect of the forcing variable is assumed to be identical and linear on both sides of the cut-off point.

Identification

In the Netherlands, as in most western societies, primary school children's grade level is strongly associated with their age. Until 1985 an official law in the Netherlands determined that children who were born before the first of October of a given year would progress from kindergarten into first grade, while children born after this cut-off date would remain in kindergarten one more academic year. Although this rule is no longer enforced today, schools still have a strong tendency to adhere to this guideline (Rijksoverheid, 2013b). This makes age a suitable forcing variable.

Nevertheless, children repeating or skipping a grade is a common phenomenon in the Netherlands, leading to misassignment of children to the control and treatment groups, resulting in biased RD estimates. These children are believed to be structurally different on background characteristics such as IQ, parental involvement in children's education and month born, compared to children with a standard school career (Corman, 2003). Therefore, we decided to remove grade repeaters and grade skippers from the analysis. This led to the exclusion of 25% of the original sample. An implication is that the application of the results of the present study is limited to children around the cut-off point with a standard school career.

Children are assigned to a treatment and a control group based on their value on the forcing variable age, measured in months. For each grade cohort included, a cut-off point is determined by the researchers. The specific process is best explained by the use of the 2nd to 3rd grade cohort as an example. First, children's age is listed for all children attending grade 2 or grade 3, which reveals that the oldest children in grade 2 at time of testing have the same age as the youngest children in grade 3. This can be attributed to the fact that compared to most previous studies investigating the effect of education on academic skills in primary school, children in the present sample are not all tested at one point in time. Instead, they are tested across a six month period, whereby the order was determined randomly by the researchers. Hence, exogenous variation in age is generated not only by differences in birth month, but also by date of test administration. Since the overlapping values observed for the age variable are a result of the testing procedure, we decide to drop the children from grade 2 who had the same age as children attending grade 3. Thus, grade 2 children who are 101 months old or older are dropped from further analyses. Subsequently, the cut-off point for the

2nd to 3rd grade cohort is defined as 101 months. Children having an age equivalent to the cut-off point or higher receive the treatment, thus attend a higher grade level and have received about a year of additional education. Next, these steps are repeated for the remaining grade cohorts. The cut-off points for the cohorts' 3rd to 4th grade, 4th to 5th grade and 5th to 6th grade are 113 months, 124 months and 136 months, respectively. To check for disproportional sorting towards the treatment group, we checked the density jump at the cut-off point. However, no jump in observations was observed at the cut-off point, indicating that children were unable to control their score in order to receive treatment.

The treatment will be estimated using a bandwith of three months around the cut-off point. This bandwith is chosen because for most grade cohorts the regression lines could be adequately modeled by a linear functional form and the control and treatment group were equal in expectations. The latter is concluded after checking the comparability of the control and treatment group per grade cohort on several observable background characteristics. The results are presented in Table 2. Continuous variables are analyzed using a T-test. A positive t-test score denotes that the treatment group scores higher on the variable than the control group. A negative t-test score indicates the opposite. For the categorical variables (mother tongue, diagnosis, gender, school and region), the z-statistic of the Mann-Whitney test are presented.

Table 2 - Comparability treatment and control group

	Grade cohorts			
	gr 2-3	gr 3-4	gr 4-5	gr 5-6
	Sample size			
Control group	39	49	60	37
Treatment group	26	11	35	31
Variable	*T-test scores / Mann-Whitney test scores*┼			
IQ (standardized scores)	-1,16	1,06	1,41	-1,18
reading performance (raw scores)	1,59	3,20***	0,24	2,14**
mother tongue	0,00	-0,35	0,95	-1,59
Diagnosis	1,75*	-1,10	0,55	-1,19
Gender	0,00	0,33	-0,59	-1,85*
School	-0,26	-0,81	1,08	0,22
Region	-0,29	-1,85*	0,80	1,14
degree urbanization	-0,30	-2,16**	1,10	-0,10
size school (# of students)	-0,21	2,96***	0,51	2,45**
percentage special need students	0,21	-0,94	0,60	-2,47**

* p < .10; ** p < .05; *** p < .01

┼ continuous variables = t-test; categorical variables = Mann-Whitney test

Two grade cohorts, namely 2nd to 3rd grade and 4th to 5th grade are not significantly different on any of the analyzed covariates. But this is not the case for the remaining two grade cohorts. In cohort 3rd to 4th grade, the reading performance of the control group is significantly lower than the treatment group. In line with this, an RD estimation in which the outcome variable is replaced by the observed covariate reading performance reveals a discontinuity at the cut-off score. This discontinuity disappears if standard reading scores instead of raw reading scores are used as outcome measure. Standardized scores take factors such as age and extra schooling into account. Since reading is a skill which they learn in school, similar to math, children in lower grades generally have less experience with reading. Therefore it is not surprising that we also observe a discontinuity for this outcome variable. Moreover, the control group attends on average schools with fewer students, but a higher degree of urbanization. Analysis of the distribution of schools in the control and treatment sample shows that two schools are underrepresented in the treatment sample compared to the control sample. Both of these schools are characterized by a high degree of urbanization and average to low school size. This underrepresentation and distribution of school characteristics could explain the significant difference reported for the school specific characteristics degree of urbanization and school size in cohort 3rd to 4th grade. Lastly, with respect to this grade cohort it should be noted that the distribution of the children around the cut-off is strongly disproportional. The control and treatment group consists out of 49 and 11 children, respectively. Similar to cohort 3rd to 4th grade, the control children in cohort 5th to 6th grade score significantly lower on the reading performance measure and they attend on average schools with fewer students. In addition, control children attend on average schools with a higher percentage of special need students. Like before, we observe that the significant difference in reading performance between control and treatment group disappears when using standardized instead of raw scores. The significant difference on school specific characteristics might again be the result of an unequal distribution of the schools in the control and treatment group. Despite being less extensive than in cohort 3rd to 4th grade, it is still present. Thus, the finding that cohorts' 3rd to 4th grade and 5th to 6th grade are unequal in expectations implies that the results of the simple model specified in equation 1 should be interpreted with caution in these cohorts. This puts more emphasis on the importance of using the extended model, represented in Equation 3 (see below). Furthermore, it is important to note that the few significant differences we find for bandwidth 3 are much smaller differences that are less significant compared to the full sample of students in those two grades.

Before proceeding with the RD analyses, the mean scores of the control group and experimental group on the three mathematics outcome measures are compared for each grade cohort. The results are shown in Table 3. Overall, the experimental group performed better on the tasks than the control group. Moreover, children's mean achievement increased across grade levels. To avoid confusion, it is worth remarking that for the task number line placement, a higher score expresses a lower performance, as we calculated absolute error scores to quantify children's achievement on this measure. Striking is the finding that the mean difference between the control group and experimental group is insignificant for some of the conducted comparisons. In the 3rd to 4th grade cohort and the 5th to 6th grade cohort, the control groups mean score on the number line placement task is not significantly lower than the experimental groups mean score. Furthermore, no significant difference is observed for children's mean mental arithmetic skills in the 4th to 5th grade cohort and the 5th to 6th grade cohort.

Table 3 - Means of the control and experimental group per cohort on the outcome variables

		grade cohorts							
		grade 2-3		grade 3-4		grade 4-5		grade 5-6	
Outcome variable	*Control/treat-ment group*	*Mean*	*T-test (dif = mean(1) - mean(0))*	*Mean*	*T-test (dif = mean(1) - mean(0))*	*Mean*	*T-test (dif = mean(1) - mean(0))*	*Mean*	*T-test (dif = mean(1) - mean(0))*
Mental arithmetic	0	0,22	0,12***	0,46	,22**	0,66	0,05	0,80	0,04
	1	0,35		0,68		0,71		0,84	
Ordinal judgment	0	0,32	0,05**	0,37	,11*	0,43	0,14***	0,50	0,10***
	1	0,36		0,48		0,56		0,60	
Number line placement	0	15,96	-4,82	9,39	-0,24	7,51	1,89***	5,67	-0,37
	1	11,15		9,15		5,62		5,31	

* p < .10; ** p < .05; *** p < .01

In the RD analyses, two specifications are tested separately per grade cohort for each outcome measure. The first specification is a simple linear RD model, in which the outcome variable is regressed on age and grade. Inserting the predictor variables of the present study into equation 1, gives us Equation 2:

$$Y_i = \beta_0 + \beta_1 \text{ age} + \beta_2 \text{ grade} + \varepsilon_{i.} \quad (2)$$

Our primary interest is the coefficient β_2, which reflects the magnitude of the treatment effect grade. The variable age in this expression is recentered by subtracting the observed age score from the chosen cut-off point. The second specification is more complex, including the covariates general response speed, reading performance, IQ, gender, mother tongue and school fixed effects as well. It is modeled by Equation 3:

$$Y_i = \beta_0 + \beta_1 \text{ age} + \beta_2 \text{ grade} + \beta_3 X_i + \varepsilon_{i.} \quad (3)$$

Although the inclusion of covariates, denoted by X_i, should not significantly alter the estimated discontinuity, they can decrease the sampling variability and improve the estimate precision. In addition, as discussed in the introduction, previous literature has found the estimate of the education effect to be affected by the inclusion of child specific characteristics. Yet, only a limited number of studies applying a natural- or quasi-experiment design have incorporated several child specific factors and school fixed effects in their specifications when estimating the return to education.

5. Results

Regression discontinuity outcomes

The results of the different RD estimations for the three mathematics sub-skills mental arithmetic, ordinal judgment and number line placement are presented in Table 4. The columns titled M1 display the results of the specification equation 2 and the columns labeled M2 the results of the specification equation 3. Ensuing is a more detailed discussion of the observed effects per mathematics sub-skill, namely mental arithmetic skills, ordinal judgment skills and number line placement respectively. Subsequently, the magnitude of the treatment effect observed in the different tasks and grade cohorts is compared.

Having received additional education due to the enrollment in a higher grade level positively affects children's mental arithmetic skills in the youngest and oldest grade cohort. For cohort 2nd to 3rd grade, the regression coefficient in specification M2 shows that children attending the higher grade level are able to correctly answer about 0.227 multiplication questions per second more than children enrolled in the lower grade level. The results of cohort 5th to 6th grade also point towards an education effect, although the strength of the effect, as indexed by the significance level, decreases in response to the inclusion of covariates (from 0.722 to 0.434). On average, children in grade 6 outperform children in grade 5 by 0.434 multiplication questions per second. Thus, our results indicate that an effect is limited to the grade cohorts in which explicit attention is given to the practice of mental arithmetic skills and not merely implicit repetition of the subject matter. In the youngest grade cohort children have just learned the mental arithmetic tables, while in the oldest grade this sub-skill is explicitly covered again in order to prepare children for the national exit exam of primary school (CITO, 2013). Age, a proxy variable for different maturational factors, only affects children's mental arithmetic in the oldest grade cohort, having a regression coefficient of -0.178 in specification M1 and -0.119 in specification M2. The negative effect

Table 4 - Return to education on mathematics sub-skills while controlling for child-specific characteristics and school attended

		Task: Mental arithmetic								Task: Ordinal judgement								Task: Number line placement							
		grade 2-3		*grade 3-4*		*grade 4-5*		*grade 5-6*		*grade 2-3*		*grade 3-4*		*grade 4-5*		*grade 5-6*		*grade 2-3*		*grade 3-4*		*grade 4-5*		*grade 5-6*	
Predictor		*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*
grade (education)	*Coef.*	0,17**	0,23***	0,19	-0,14	-0,02	0,09	0,72**	0,43*	0,12**	0,15*	0,11	-0,05	0,13*	0,19**	0,28**	0,20*	1,83	0,27	-1,96	0,25	-2,31	-3,00	-4,83**	-3,784*
	Std Err	0,07	0,08	0,21	-0,19	-0,17	0,15	0,29	0,25	0,05	0,06	0,10	-0,11	0,07	0,07	0,11	0,11	4,60	4,50	-4,29	4,83	-1,83	-1,83	-1,98	-2,02
age (forcing variable)	*Coef.*	-0,02	-0,04	0,01	0,05	0,02	0,00	-0,18**	-0,12*	-0,02	-0,03*	0,00	0,02	0,00	-0,02	-0,05*	-0,05	-2,04	-1,48	0,48	0,32	0,12	0,24	1,17**	0,76
	Std Err	-0,02	-0,02	0,05	0,04	0,04	-0,04	-0,07	-0,06	-0,01	-0,02	0,02	0,03	0,02	-0,02	-0,03	-0,03	-1,28	-1,30	1,05	1,10	0,47	0,47	0,49	0,51
response speed	*Coef.*		-0,05		0,26*		0,14		-0,09		-0,01		0,13		0,19***		0,20**								
	Std Err		-0,08		0,13		0,11		-0,13		-0,06		0,08		0,05		0,06								
reading ability	*Coef.*		0,00		0,01***		0,01***		0,01***		0,00		0,00*		0,00		0,00		-0,05		-0,04		-0,02		-0,02
	Std Err		0,00		0,00		0,00		0,00		0,00		0,00		0,00		0,00		-0,04		-0,04		-0,02		-0,02
IQ	*Coef.*		0,00		-0,02		-0,01		-0,01		0,00		-0,00		-0,00		0,00		-0,54**		-0,15		-0,06		-0,42***
	Std Err		0,00		-0,01		-0,01		-0,02		0,00		-0,01		0,00		0,01		-0,25		-0,25		-0,12		-0,14
gender	*Coef.*		-0,02		0,01		0,06		-0,03		0,00		0,00		0,04		0,01		-5,78***		-4,44***		-2,16**		0,02
	Std Err		-0,03		0,06		0,07		-0,09		0,02		0,04		0,03		0,04		-1,75		-1,60		-0,82		0,72
mother tongue	*Coef.*		0,03		0,30**		0,04		0,02		0,05		0,09		-0,03		0,04		3,32		2,00		-0,98		-0,63
	Std Err		0,06		0,14		0,12		0,16		0,05		0,08		-0,06		0,07		3,63		3,53		-1,53		-1,30
school dum 1	*Coef.*		0,00		0,38***		0,36***		0,40**		-0,04		0,09		0,08		0,07		-1,18		0,64		-1,03		-0,45
	Std Err		0,05		0,12		0,13		0,17		-0,04		0,07		0,06		0,08		-3,25		3,13		-1,56		-1,43
school dum 2	*Coef.*		(.)		0,20		0,19		0,40**		(.)		0,06		0,00		0,11		(.)		2,88		1,82		0,02
	Std Err		(.)		0,13		0,18		0,19		(.)		0,08		0,09		0,09		(.)		3,29		2,21		1,54
school dum 3	*Coef.*		0,02		0,27**		0,47***		0,35**		-0,01		0,02		0,04		0,08		-1,77		0,66		-2,12		1,17
	Std Err		0,05		0,11		0,11		0,15		-0,04		0,07		0,05		0,07		-3,22		2,91		-1,36		1,25
school dum 4	*Coef.*		-0,07		0,14		0,09		0,10		-0,06		0,11		-0,04		0,00		0,31		-0,64		0,41		0,69
	Std Err		-0,06		0,13		0,12		0,16		-0,04		0,08		-0,06		-0,07		3,63		-3,26		1,48		1,31
school dum 5	*Coef.*		0,03		0,21*		0,09		0,48**		0,01		0,08		-0,03		0,02		0,31		1,41		1,09		0,10
	Std Err		0,06		0,12		0,11		0,19		0,04		0,07		-0,05		0,09		3,35		3,01		1,40		1,55
Constant	*Coef.*	0,19***	0,21	0,48***	-0,75	0,70***	-0,43	0,40**	-0,58	0,27***	0,16	0,37***	-0,11	0,43***	-0,03	0,40***	-0,25	12,06***	34,24***	10,32***	18,31*	7,77***	15,03***	8,29***	23,32***
	Std Err	0,04	0,15	0,11	-0,39	0,10	-0,36	0,17	-0,65	0,03	0,11	0,05	-0,23	0,04	-0,17	0,07	-0,30	2,75	7,90	2,22	9,02	1,10	4,26	1,18	5,21
R-squared		0,20	0,19	0,05	0,45	-0,02	0,34	0,06	0,41	0,07	0,04	0,05	0,16	0,13	0,27	0,09	0,23	0,10	*0,27*	-0,03	0,08	0,03	0,14	0,06	0,13
N		65	65	60	60	94	92	68	68	65	65	60	60	94	92	68	68	63	63	60	60	94	92	68	68
p		0,00	0,02	0,10	0,00	0,76	0,00	0,05	0,00	0,03	0,30	0,08	0,06	0,00	0,00	0,02	0,01	0,02	0,00	0,89	0,16	0,08	0,01	0,06	0,06

* p < .10; ** p < .05; *** p < .01; (.) = no cases presented in cohort

of age seems counterintuitive. Yet, the fact that we already control for grade and that we use a relatively small sample and a small age range in which age differences are measured in months, could have made the variable age prone to random effects. Regarding the covariates, school fixed effects has the most noticeable effect, ranging from 0.206 in cohort 3rd to 4th grade to 0.476 in cohort 5th to 6th grade. On average children attending schools adhering to a more traditional classroom teaching style have higher mental multiplication skills than similar children attending the reference school. These findings imply that attending a school emphasizing a natural learning approach and multigrade teaching negatively affects the development of children's mental arithmetic skills. However, no significant school fixed effect is found in cohort 2nd to 3rd grade, when children first encounter mental arithmetic's. This suggests that not differences in the approach to the initial learning, but rather the form and/ or extend to which the arithmetic skills are reinforced in higher grades determines variation in children's mental arithmetic proficiency between the more traditional schools and the reference school. Which factor exactly causes the reference school children to perform lower can only be speculated about.

The fact that in the reference school multiple grade levels with different expertise are taught simultaneously within one classroom requires teachers to differentiate and can negatively influence the attention and time a teacher has to repeat previously learned mathematics skills. Moreover, this school emphasizes natural learning. One can question whether natural learning is a suitable approach to stimulate the sustained practice of multiplication skills. Geary (1995) postulated that multiplication is an ability which not only requires sustained experience with the content through formal schooling, but that children's biological motivation to engage in this effortful activity is rather low. Hence, children might not be sufficiently internally motivated to progressively work on their arithmetic skills if learning is child-centered. Lastly, another covariate which has a significant effect on mental arithmetic proficiency is children's reading ability. Though the regression coefficients are small, ranging from 0.006 to 0.010, its impact size is considerable as expressed by standardized coefficients ranging from 0.357 to 0.630 of a standard deviation. A possible explanation of this significant effect could be that in Western societies multiplication is mostly taught by verbal transmission and therefore children's verbal skills might be important to the learning of arithmetic facts. Furthermore, in the Netherlands children often learn numerical combinations by story problems, which require sufficient reading ability.

A positive education effect was also found on the ordinal judgment task for three out of four grade cohorts, indicating that an additional year of education leads to significantly higher ordinal judgment skills in children across multiple consecutive grade levels in primary school. In cohort 3rd to 4th grade no significant education effect was observed. We cannot think of a theoretical reason which could account for this finding. Instead, we can claim that the absence of an effect can probably be attributed to the strongly disproportionate distribution of individuals across the treatment (N = 11) and control group (N = 49). Like for mental arithmetic, only a limited effect of age is identified. Particularly, a one month increase in age at testing moment, leads to a reduction in children's ordinality judgment scores of 0.030 and 0.048 in specification M2 for cohort 2nd to 3rd grade and specification M1 for cohort 5th to 6th grade, respectively, given that we have already controlled for grade. In both grade cohorts the age effect was not consistent across the specifications. Children's ordinal judgment performance is generally not sensitive to individual differences in covariates, except for general response speed. A faster response speed of one unit significantly improves children's ordinal judgment skills by 0.192 units and 0.200 units in cohort 4th to 5th grade and 5th to 6th grade, respectively. Yet, this is not surprising because the ordinality task is a timed task requiring the fact processing of ordinal number information.

Contrary to expectation, no education effect on children's number line placement skills is observed for the younger grade cohorts, but only in oldest cohort 5th to 6th grade. This is surprising, since one of the key goals in grade 3 is the learning of the number structure from 0-1000, while in the lower grade levels children should be able to place numbers from 0-100 on a number line. It is possible that an effect has gone unnoticed due to the small sample size of the present study. However, children's number line placement skills might also be determined more by other factors than education. Accordingly, we found a significant effect of gender and nonverbal IQ on number line placement skills. In cohort 2nd to 3rd grade, boys make on average 5.781 fewer errors when placing numbers from 0-1000 on a number line than girls. Note that number line placement skills are indexed by absolute error scores and therefore a negative sign denotes better performance. This finding is in line with previous studies who reported a link between children's gender and estimation abilities, indexed by for example number line tasks (Hannula, 2003; van den Heuvel-Panhuizen, 2004). Concerning children's nonverbal IQ level, a significant effect is observed in cohort 2nd to 3rd grade and 5th to 6th grade, in which a one point increase in IQ score on a scale of 0 to 36 corresponds to a decrease in absolute errors of 0.538 and 0.417, respectively. Another explanation regarding the absence of a significant education effect on number line placement skills in the younger grade cohorts needs to be considered as well. It could simply be that the manner in which teachers transmit number line placement knowledge is not making optimal use of they way in which children acquire this skill best. Unfortunately, the paper at hand cannot distinguish among the discussed accounts.

To summarize, an education effect captured by the variable grade, is mainly observed for two mathematics sub-skills, particularly mental arithmetic's and ordinal judgment in cohorts 2nd to 3rd grade and 5th to 6th grade. As is displayed in Table 4, the regression coefficients in the older grade cohort are considerable larger than in the younger grade cohort. This suggests that teaching mental arithmetic to older children is more effective, in the sense that their learning curve is steeper. However, this finding could also be attributable to the fact that differences between student performances in this age group are substantially higher than in the lower age groups. Therefore, standardized regression coefficients are computed as well, which quantify the strength of an effect in standard deviations and consequently allow for a better comparison of the magnitude of the education effect between grade cohorts and mathematics sub-skills. The results are presented in Table 5.

Table 5 - Standardized coefficients of the treatment variable grade for each task, specification and grade cohort

		Grade cohorts							
		grade 2-3		grade 3-4		grade 4-5		grade 5-6	
Task		*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*
Mental arithmetic	*Beta*	0,65	0,85	0,24	-0,17	-0,02	0,12	0,87	0,52
Ordinal judgment	*Beta*	0,67	0,83	0,29	-0,12	0,36	0,53	0,83	0,60
Number line placement	*Beta*	0,12	0,02	-0,12	0,02	-0,28	-0,36	-0,85	-0,67

* p < .05; ** p < .01; *** p < .001

The size of the education effect on the two mathematics sub-skills mental arithmetic's and ordinal judgment is substantial for the youngest and oldest grade cohort, ranging from 0.522 to 0.867. For example, in cohort 2nd to 3rd grade in specification M2, children in the higher grade score 0.845 standard deviations higher on the mental arithmetic task than children in the lower grade level. The difference in the amount of formal education received by the control and treatment group is considerable, adding up to approximately one year, and for that reason the relatively large magnitude of the education effect is not unexpected. Previous studies have reported substantial education effects in the lower grade levels as well (Cahan & Davis, 1987; Luyten, 2006; Luyten, Tymms, & Jones, 2009). Contributing to this previous knowledge is the present result that the impact is high in the oldest grade cohort as well. This implies that not only the initial learning, but also the explicit repetition of basic mathematics' skills is crucial to the continual development of mental arithmetic and ordinal judgment skills.

Robustness analysis

To verify the robustness of the results, we run three additional RD analyses with slight alterations. These RD analyses are conducted for both specifications modeled previously and all three mathematic outcome measures. The results are shown in Table 6. First, a bandwith of four instead of three months was used. Several of the education effects reported previously are now observed as well, though the impact decreases. For example, the regression coefficient capturing the education effect on children's mental arithmetic skills in cohort 2nd to 3rd grade declines from 0.227 to 0.182 for the specification M2. A significant but also reduced education effect is found on children's ordinal judgment skills in cohort 4th to 5th grade and 5th to 6th grade. Yet, not all education effects remain significant. An education effect is no longer observed on children's mental arithmetic skills in cohort 5th to 6th grade, their ordinal judgment skills in cohort 2nd to 3rd grade and their number line placement skills in cohort 5th to 6th grade. Note that the extension of the bandwith to 4 months results in similar comparability between children of the control and treatment group. However, one might question whether the assumption of the linear functional form can still be made when increasing the bandwidth to 4 months; this might explain the disappearance of some of the education effects which were found when analyses were restricted to a bandwith of three months.

Afterwards, an RD is implemented, but instead of raw scores the log transformed variables were included. Most of the significant education effects were replicated, though the strength as measured by the significance level decreased slightly for many of the effects. For example, in the cohort 2nd to 3rd grade the significance level of the education effect on mental arithmetic in specifications M2 decreases from $p < .01$ to $p < .05$. However, for cohort 5th to 6th grade, education no longer has a significant impact on children's mental arithmetic skills and their number line placement skills in specification M2.

Lastly, an RD including the children which were earlier removed from the sample because they were misassigned to the control or treatment group based on their score on the forcing variable is conducted. Notice that the children who repeated or skipped a class are still excluded from the analyses. We called this RD approach a fuzzy RD design. All education effects are reproduced, except for the effect on mental arithmetic skills in

Table 6 - Robustness checks

		Task: Mental arithmetic								Task: Ordinal judgment								Task: Number line placement							
		grade 2-3		*grade 3-4*		*grade 4-5*		*grade 5-6*		*grade 2-3*		*grade 3-4*		*grade 4-5*		*grade 5-6*		*grade 2-3*		*grade 3-4*		*grade 4-5*		*grade 5-6*	
Predictor		*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*	*M1*	*M2*
Robustness check: bandwith 4																									
grade (education)	*Coef.*	0,17**	0,18**	0,15	-0,09	-0,03	0,04	0,30	0,17	0,06	0,08	0,11	-0,02	0,13*	0,15**	0,17*	0,12	-0,48	-1,12	3,09	5,64	-2,29	-2,97*	-1,39	-2,22
	Std Err	0,07	0,07	0,16	-0,15	-0,15	0,13	0,21	0,19	0,05	0,05	0,07	-0,08	0,07	0,06	0,09	0,09	-3,40	-3,04	3,39	3,51	-1,72	-1,70	-2,01	-1,98
R-squared		0,14	0,24	0,04	0,31	0,00	0,31	0,00	0,31	0,07	0,09	0,02	0,18	0,11	0,22	0,08	0,19	0,16	0,37	0,01	0,17	0,02	0,12	-0,01	0,15
N		107	107	91	91	119	117	103	103	107	107	92	92	119	117	103	103	102	102	92	92	119	117	103	103
p		0,00	0,00	0,06	0,00	0,42	0,00	0,36	0,00	0,01	0,04	0,16	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,29	0,01	0,14	0,01	0,48	0,01
Robustness check: log independent and outcome variables																									
grade (education)	*Coef.*	0,50*	0,69**	0,25	-0,49	0,10	0,25	0,81**	0,34	0,45**	0,53***	0,33	-0,11	0,26	0,33*	0,52**	0,43*	0,26	0,15	-0,30	-0,09	-0,23	-0,31	-0,49*	-0,31
	Std Err	0,25	0,26	0,42	-0,37	0,26	0,22	0,39	0,32	0,17	0,19	0,29	-0,34	0,16	0,17	0,24	0,25	0,36	0,35	-0,39	-0,42	-0,21	-0,20	-0,27	-0,28
R-squared		0,23	0,29	0,01	0,45	-0,01	0,39	0,03	0,43	0,09	0,05	0,02	0,06	0,12	0,13	0,08	0,10	0,10	0,25	-0,02	0,16	0,03	0,19	0,02	0,08
N		65	65	60	60	94	92	68	68	65	65	60	60	94	92	68	68	63	63	60	60	94	92	68	68
p		0,00	0,00	0,27	0,00	0,63	0,00	0,12	0,00	0,02	0,25	0,21	0,22	0,00	0,02	0,02	0,11	0,02	0,00	0,69	0,05	0,10	0,00	0,18	0,15
Robustness check: fuzzy design																									
grade (education)	*Coef.*	0,13***	0,11**	0,22	0,04	-0,07	-0,03	0,38	0,35	0,09***	0,09***	0,15**	0,06	0,13***	0,14***	0,17*	0,10	-5,28**	-4,00	-1,53	-0,39	-1,47	-1,91**	-4,29**	-2,58
	Std Err	0,04	0,05	0,16	0,13	-0,09	-0,08	0,24	0,24	0,03	0,03	0,08	0,08	0,03	0,04	0,09	0,11	-2,55	-2,43	-3,28	-3,34	-0,92	-0,95	-1,69	-1,97
R-squared		0,13	0,15	0,05	0,46	0,01	0,33	0,01	0,34	0,12	0,08	0,07	0,15	0,11	0,19	0,05	0,19	0,07	0,23	-0,03	0,12	0,02	0,09	0,06	0,14
N		82	82	68	68	140	138	70	70	82	82	68	68	140	138	70	70	78	78	68	68	140	138	70	70
p		0,00	0,02	0,08	0,00	0,23	0,00	0,25	0,00	0,00	0,12	0,04	0,04	0,00	0,00	0,06	0,02	0,03	0,00	0,88	0,07	0,08	0,02	0,04	0,04

* p < .05; ** p < .01; *** p < .001

specification M1 and M2 and on number line placement in specification M2 in cohort 5^{th} to 6^{th} grade. In addition, in two models a previously undetected education effect was now observed. Using the fuzzy RD design we found that third grade children scored significantly higher on the ordinal judgment task in specification M2 and on the number line placement task in specification M1 than the second grade children. We conclude that overall the education effects are robust for the cohort 2^{nd} to 3^{rd} grade and to a great extent also for the cohort 5^{th} to 6^{th} grade. The overall conclusion that the education effect is strongest in the youngest and oldest grade cohort is confirmed in these robustness analyses.

6. Conclusions

Children's progression in mathematics skills is said to be insufficient in order to cope with an increasing quantitative world in the future. As more advanced skills built upon more elementary mathematics skills, it is important to know how effective early education is in fostering children's proficiency on basic mathematic sub-skills. Previous studies have examined this issue, but they frequently used general mathematics achievement tests and did not investigate the consistency of the estimates across multiple primary school grade levels (e.g. Cahan & Davis, 1987; Cliffordson, 2010). Hence, the paper at hand adds to the literature by exploring the variability in the effect of an additional year of education on three elementary mathematical skills across different grade levels.

The RD results indicate that the effect of education is contingent upon the grade cohort and mathematic sub-skill analyzed. The effect is most pronounced for the mental arithmetic task and ordinal judgment measure in two cohorts, particularly the 2^{nd} to 3^{rd} grade and the 5^{th} to 6^{th} grade cohorts. In the younger grade cohort children are being taught these mathematics sub-skills, while in the older grade cohort children explicitly repeat these skills in class to prepare for a national exit exam. Standardized coefficients show that the magnitude of the education effect was substantial for both of these grade cohorts. On the practical side, these findings show that education is not only important as a mechanism through which children learn new mathematics skills, but is equally crucial to the reinforcement and improvement of these skills. The failure to find strong education effects for the cohorts 3^{rd} to 4^{th} grade and 4^{th} to 5^{th} grade demonstrates that advancement in mathematics skills across grades is not linearly attributable to an increase in education. Instead, in the absence of explicit practice of mathematics sub-skills in the classroom, the progress observed between the lower and higher grade level is mostly driven by other aspects, such as maturational and school factors. Hence, we suggest that introducing substantially more explicit and frequent repetition of the mathematics sub-skills in the in between grade levels has the highest potency of developing children's skills to a higher level.

The relevance of regular, explicit repetition of elementary mathematics sub-skills in primary school is further supported by the fact that being enrolled in the reference school in which child centered, natural learning and multi grade teaching is emphasized, has a negative impact on children's mental arithmetic proficiency. Children do not seem to have an internal motivation to keep practicing their arithmetic skills (Geary, 1995) and therefore repeated explicit instruction is necessary for the improvement of children's mental arithmetic proficiency. However, it should be noted that this implication seems to be specific to the sub-skill mental arithmetics as no school effects were found to have a significant effect on the other mathematics sub-skills assessed.

The expected effect of education on children's number line placement skills in the cohort 2^{nd} to 3^{rd} grade is not found, despite that this skill is emphasized in grade 3. Possible explanations include ineffective education, unnoticeable effect due to the small sample size and other factors more important. In line with the latter is the present result that children's number line placement skills is significantly affected by nonverbal IQ and gender in the majority of the grade cohorts.

The present study also has scientific implications. First, estimates of the education effect on mathematics skills operationalized by scores on a general mathematics achievement test cannot simply be generalized to individual mathematic sub-skills. As our study reveals, the magnitude of the education effect is dependent upon the mathematics sub-skill assessed. Second, education effect estimates are sensitive to the grade levels assessed. Consequently, education effect estimates should not be generalized to other grades than the ones assessed.

Lastly, future research addressing the impact of education on mathematics sub-skills in primary school grades should take a couple of limitations of the study at hand into account in order to obtain better education effect estimates. Although the use of a cross-sectional data set collected for the purpose of a test standardization study allowed us to include child-specific characteristics in our specifications, it also meant that only a relatively small sample was incorporated in each grade cohort. Furthermore, we had no information on teacher-specific characteristics or on how many hours of classroom mathematics education children received exactly. Our attempt to gather this information was unsuccessful, unfortunately, as only a very small share of the teachers returned the teacher questionnaire.

References

Angrist, J. D., & Pischke, J.-S. (2009). *Harmless econometrics: An empiricist's companion.* Oxfordshire, UK: Princeton University Press

Aster, M. G. von, & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine and Child Neurology, 49,* 868-873. doi:10.1111/j.1469-8749.2007.00868.x

Bernstein, D. A., Penner, L. A., Clarke-Stewart, A., & Roy, E. J. (2006). *Psychology.* Boston, MA: Houghton Mifflin Company

Bisanz, J., Morrison, F. J., & Dunn, M. (1995). Effects of age and schooling on the acquisition of elementary quantitative skills. *Developmental Psychology, 31*, 221-236. doi:10.1037/0012-1649.31.2.221

Cahan, S., & Davis, D. (1987). A between-grade-levels approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal, 24,* 1-12. doi:10.3102/00028312024001001

Cito. (2013). *Eindtoets Basisonderwijs.* Retrieved from http://www.cito.nl/onderwijs/primair%20onderwijs/cito_volgsysteem_po/eindtoets_basisonderwijs/747b121dd9444486b43117c5bac96bb0.aspx

Cliffordson, C. (2010). Methodological issues in investigations of the relative effects of schooling and age on school performance: the between-grade regression discontinuity design applied to Swedish TIMSS 1995 data. *Educational Research and Evaluation: An International Journal on Theory and Practice*, *16*, 39-52. doi:10.1080/13803611003694391

Cook, T. (2008). Waiting for life to arrive: A history of the regression-discontinuity design in psychology, statistics, and economics. *Journal of Econometrics*, *142*, 636-654. doi:10.1016/j.jeconom.2007.05.002

Corman, H. (2003). The effects of state policies, individual characteristics, family characteristics, and neighbourhood characteristics on grade repetition in the United States. *Economics of Education Review*, *22*, 409-420. doi:10.1016/S0272-7757(02)00070-5

Crawford, C., & Cribb, J. (2013). *Reading and maths skills at age 10 and earnings in later life: a brief analysis using the British Cohort Study* (CAYT impact study: REP03). Retrieved from Institute for Fiscal Studies, Centre for Analysis of Youth Transitions website: http://www.ifs.org.uk/publications/6624

Dougherty, C. (2003). Numeracy, literacy and earnings: Evidence from the National Longitudinal Survey of Youth. *Economics of Education Review, 22,* 511 – 521. doi:10.1016/S0272-7757(03)00040-2

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2005). First grade and educational attainment by age 22: A new story. *American Journal of Sociology, 110,* 1458-1502. doi:10.1086/428444

Fitzpatrick, M. D., Grissmer, D., & Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review, 30,* 269-279. doi:10.1016/j.econedurev.2010.09.004

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist, 50,* 24-37. doi: ???

Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal, 47,* 434-465. doi:10.3102/0002831209350494

Harmon, C., Oosterbeek, H., & Walker, I. (2003). The returns to education: microeconomics. *Journal of Economic Surveys, 17,* 115-156. doi:10.1111/1467-6419.00191

Hannula, M. S. (2003). Locating fractions on a number line. In N A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th Conference of PME-NA* (Vol. 3, pp. 3-24). Honolulu, HI: PME.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 21*, 377-408. doi:10.1080/09243453.2010.500097

Heckman, J., & Vytlacil, E. (2001). Identifying the role of cognitive ability in explaining the level of and change in the return to schooling. *The Review of Economics and Statistics, 83,* 1-12. doi:10.1162/003465301750159993

Huang, J., Maassen van den Brink, H., & Groot, W. (2009). A meta-analysis of the effect of education on social capital. *Economics of Education Review, 28,* 454-464. doi:10.1016/j.econedurev.2008.03.004

Inspectie van het Onderwijs. (2012). *Onderzoek in het kader van het vierjaarlijks bezoek bij de J.H. Snijdersschool* (Rapport 126659). Retrieved from Ministerie van Onderwijs, Cultuur en Wetenschap website: http://tkrtp.owinsp.nl/brincode/13RY/zoekresultaat?xl=1&p1=brin&p2=%3D&p3=13RY&p1=%23&p2=curpg&p3=1&sector=PO&sch_id=911306.22&arr_id=218263

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics, 86,* 226-244. doi:10.1162/003465304323023778

Jacob, S. N., & Nieder, A. (2008). The ABC of cardinal and ordinal number representations. *Trends in Cognitive Sciences, 12,* 41-43. doi:10.1016/j.tics.2007.11.006

J. H. Snijdersschool. (2013). *Schoolgids 2012-2013.* Retrieved from http://www.natuurlijksnijdersschool.nl/files/schoolgids/Snijdersschool_Web_Single.pdf

Lu, L., Weber, H. S., & Spinath, F. M. (2011). Predicting school achievement from cognitive and non-cognitive variables in a Chinese sample of elementary school children. *Intelligence, 39,* 130-140. doi:10.1016/j.intell.2011.02.002

Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: Regression-discontinuity applied to TIMSS-95. *Oxford Review of Education, 32,* 397-429. doi:10.1080/03054980600776589

Luyten, H., Tymms, P., & Jones, P. (2009). Assessing school effects without controlling for prior achievement? *School Effectiveness and School Improvement, 20,* 145-165. doi:10.1080/09243450902879779

Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review, 26,* 629-640. doi:10.1016/j.econedurev.2006.08.001

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics, 77,* 251-266. Retrieved from http://www.mitpressjournals.org/loi/rest

Naito, M., & Miura, H. (2001). Japanese children's numerical competencies: Age- and schooling-related influences on the development of number concepts and addition skills. *Developmental Psychology, 37,* 217-230. doi:10.1037/0012-1649.37.2.217

Nichols, A. (2009). *Causal inference with observational data: Regression discontinuity and related methods in stata.* Retrieved from http://www.stata.com/meeting/germany09/nichols.pdf

Raven, J., Court, J. H., & Raven, J. C. (1995). *Coloured progressive matrices.* Oxford, UK: Oxford Psychologists Press

Rijksoverheid (2013a). *Kerndoelen basisonderwijs.* Retrieved from http://www.rijksoverheid.nl/onderwerpen/basisonderwijs/kerndoelen-basisonderwijs

Rijksoverheid (2013b). *Wat is de leerplicht?.* Retrieved from http://www.rijksoverheid.nl/onderwerpen/leerplicht/vraag-en-antwoord/wat-is-de-leerplicht.html

Ritchie, S. J., & Bates, T. C. (in press). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science.* doi:10.1177/0956797612466268

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). *Standards for regression discontinuity designs.* Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_rd.pdf

Sims, D. P. (2008). Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review, 27,* 58-68. doi:10.1016/j.econedurev.2006.05.003

Spinath, B., Freudenthaler, H. H., & Neubauer, A. C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Personality and Individual Differences, 48,* 481-486. doi:10.1016/j.paid.2009.11.028

Van Groenestijn, M. (2008). Dyscalculie: een probleem van het kind of van het onderwijs? In M. Dolk & M. van Groenestijn (Eds.), *Dyscalculie in discussie* (pp. 64-70). Assen, Netherlands: Koninklijke Van Gorcum.

Van den Heuvel-Panhuizen, M. (2004). Girls' and boys' problems: Gender differences in solving problems in primary school mathematics in the Netherlands. In B. Clarke, D.M. Clarke, G. Emanuelsson, B. Johansson, D.V. Lambdin, F.K. Lester, A. Wallby, & K. Wallby (Eds.), *International Perspectives on Learning and Teaching Mathematics* (pp. 237-252). Göteborg, Sweden: National Center for Mathematics Education.

Van der Stap, M. (2012). *Van kerndoel tot leerlijn: Concretisering van de kerndoelen voor het speciaal onderwijs.* Amsterdam, Netherlands: SWP

TIER WORKING PAPER SERIES
TIER WP 13/12

ISBN 978-94-003-0063-7